Genome-Wide Mutational Signatures of Aristolochic Acid and Its Application as a Screening Tool

Song Ling Poon,1,2,3*, See-Tong Pang,3,4† John R. McPherson,2,5 Willie Yu,1,2,6 Kie Kyon Huang,2,5 Peiyong Guan,2,5 Wen-Hui Weng,7 Ee Yan Siew,1,2 Yujing Liu,2,8 Hong Lee Heng,1,2 Soo Ching Chong,1,2 Anna Gan,1,2 Su Ting Tay,9 Weng Khong Lim,1,2 Ioana Cutcutache,2 Dachuan Huang,1,2,7 Ruo Liang Dee Ler,1,2,4 Maarja-Lisa Naismägi,2,2 Ming Hui Lee,9 Ying-Hsu Chang,3 Kai-Jie Yu,3 Waraporn Chan-on,1,2 Bin-Kui Li,10 Yun-Fei Yuan,11 Chao-Nan Qian,11 Kwei-Fong Ng,12 Ching-Fang Wu,13 Cheng-Lung Hsu,14 Ralph M. Bunte,15 Michael R. Stratton,16 P. Andrew Futreal,17 Wing-Kin Sung,6,18 Cheng-Keng Chuang,3 Choon Kiat Ong,1,2 Steven G. Rozen,2,19† Patrick Tan,2,5,9,18† Bin Tean Teh,1,2,5†

Aristolochic acid (AA), a natural product of Aristolochia plants found in herbal remedies and health supplements, is a group 1 carcinogen that can cause nephrotoxicity and upper urinary tract urothelial cell carcinoma (UTUC). Whole-genome and exome analysis of nine AA-associated UTUCs revealed a strikingly high somatic mutation rate (150 mutations/Mb), exceeding smoking-associated lung cancer (8 mutations/Mb) and ultraviolet radiation–associated melanoma (111 mutations/Mb). The AA-UTUC mutational signature was characterized by A:T to T:A transversions at the sequence motif A[C|T]AGG, located primarily on nontranscribed strands. AA-induced mutations were also significantly enriched at splice sites, suggesting a role for splice-site mutations in UTUC pathogenesis. RNA sequencing of AA-UTUC confirmed a general up-regulation of nonsense-mediated decay machinery components and aberrant splicing events associated with splice-site mutations. We observed a high frequency of somatic mutations in chromatin modifiers, particularly KDM6A, in AA-UTUC, demonstrated the sufficiency of AA to induce renal dysplasia in mice, and reproduced the AA mutational signature in experimentally treated human renal tubular cells. Finally, exploring other malignancies that were not known to be associated with AA, we screened 93 hepatocellular carcinoma genomes/exomes and identified AA-like mutational signatures in 11. Our study highlights an unusual genome-wide AA mutational signature and the potential use of mutation signatures as “molecular fingerprints” for investigating high-throughput cancer genome data to infer previous carcinogen exposures.

INTRODUCTION

Unraveling the specific mutational effects of environmental carcinogens on the human genome is a cornerstone of cancer prevention research. Cancer genome analysis has revealed that different carcinogens, including pathogens (1, 2), cigarette smoke (3), and ultraviolet (UV) radiation (4), are often associated with distinct mutational patterns (“mutational signatures”), acting as telltale genetic fingerprints of previous exposures. Analysis of the spectra and patterns of somatic mutations in cancer genomes can also lead to a greater understanding of the mutational events triggering clonal outgrowth.

Aristolochic acid (AA) is a natural compound found in many plants of the Aristolochia genus. Aristolochia plants are commonly used in traditional herbal preparations as health supplements and remedies for various health problems including weight loss, menstrual symptoms, and rheumatism (5, 6). In the 1990s, epidemiological studies showed that AA exposure was associated with a high risk of nephrotoxicity and upper urinary tract urothelial cell carcinoma (UTUC) (7–10) caused by the ability of AA to bind DNA-forming DNA adducts (11). These findings consequently led to bans on the use of Aristolochia-containing herbal preparations in Europe and North America since 2001 and in

1NCS-VIMI Translational Research Laboratory, Division of Medical Sciences, National Cancer Centre Singapore, 11 Hospital Drive, Singapore 169610, Singapore. 2Program in Cancer and Stem Cell Biology, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857, Singapore. 3Division of Urology, Department of Surgery, Chang Gung Memorial Hospital, Linkou; School of Medicine, Chang Gung University, 5, Fusing Street, Guishan Township, Taoyuan County 333, Taiwan. 4National University of Singapore Graduate School for Integrative Sciences and Engineering, 28 Medical Drive, Singapore 117456, Singapore. 5Cancer Science Institute of Singapore, National University of Singapore, Singapore; Centre for Life Sciences, 28 Medical Drive, Singapore 117456, Singapore. 6School of Computing, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077, Singapore. 7Department of Chemical Engineering and Biotechnology, Graduate Institute of Biotechnology, National Tsing Hsiao East Road, Taipei 10608, Taiwan. 8Computational and Systems Biology, Singapore-MIT Alliance, National University of Singapore, Singapore 119077, Singapore. 9Division of Cellular and Molecular Research, National Cancer Centre Singapore, 11 Hospital Drive, Singapore 169610, Singapore. 10Department of Hepato-Biliary Oncology, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong 510060, China. 11State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong 510060, China. 12Department of Pathology, Chang Gung Memorial Hospital, Linkou; School of Medicine Chang Gung University, Taoyuan County 333, Taiwan. 13Division of Urology, Department of Surgery, Chang Gung Memorial Hospital, Chiai; School of Traditional Chinese Medicine, Chang Gung University, Taoyuan County 333, Taiwan. 14Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, Taoyuan County 333, Taiwan. 15Office of Research, Duke-NUS Graduate Medical School, Singapore 169857, Singapore. 16Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. 17Genomic Medicine, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1954, Houston, TX 77030, USA. 18Genome Institute of Singapore, 60 Biopolis Street Genome, Singapore 138672, Singapore. 19Department of Psychiatry and Behavioral Sciences, Duke-University Medical Center, Durham, NC 27710, USA.

*These authors contributed equally to this work.†Corresponding author. E-mail: teh.bin.tean@singhealth.com.sg (B.T.T.); gmstanp@duke-nus.edu.sg (P.T.); steve.rozen@duke-nus.edu.sg (S.G.R.); jacobpang@adm.cgmh.org.tw (S.-T.P.)

Asia since 2003 (9). Currently, AA is classified as a group 1 human carcinogen in the IARC Monograph (12). In Asia, AA-associated DNA adducts can be detected in the renal cortex of more than 50% of UTUC patients in Taiwan (13), and the incidence of UTUC in Taiwan (30% of urothelial cancer) is strikingly higher than that in the West (3% of urothelial cancer) (14), consistent with AA playing a major carcinogenic role in Asian UTUC. However, the detailed mechanisms by which AA-induced mutations contribute to UTUC tumorigenesis remain largely unknown, and the involvement of AA in other cancer types has also not been extensively studied. Indeed, whereas mutagens such as AA clearly act on the entire genome, to date only mutations and mutational signatures in TP53 alone have been investigated in AA-UTUC (13, 15–17). Thus, a genome-wide approach is needed to understand the full implications of AA-induced mutagenesis and how it contributes to UTUC.

Here, by whole-genome and exome sequencing, we studied a panel of UTUC patients who were exposed to AA. We identified (i) a strikingly high prevalence of somatic mutations, (ii) a specific AA mutational signature, and (iii) recurrent mutations in UTUC, notably involving chromatin-modifying genes. We recapitulated AA toxicity in vivo and the AA mutational signature in vitro AA-treated cells. Finally, we used the AA mutational signature to screen for and detect AA-like mutational patterns in hepatocellular carcinoma (HCC), which suggests an underappreciated carcinogenic role for AA (or an AA-like compound) in a subset of liver cancers.

RESULTS

AA is a highly mutagenic group 1 carcinogen

To investigate the molecular effects of AA on a genomic scale, we performed whole-genome sequencing (WGS) of an AA-UTUC (tumor 9T) and nonmalignant kidney tissue from the same patient (table S1). Clinical exposure to AA was inferred from patient medical records and case histories, including female gender, history of herbal remedy use, and compromised renal function (18, 19). Female gender is relevant because it appears that, in Taiwan, AA-containing preparations are mostly used by women for weight loss (16). The UTUC and matched normal genomes were sequenced to 33X mean haploid genome coverage (table S2). Genome Analyzer Toolkit (GATK) software was used to identify single-nucleotide variants as previously described (1, 2). To identify somatic mutations, we excluded from analysis germ-line variants catalogued in either the dbSNP135 or 1000 Genome Project database and then subtracted the sequence variants of the normal genomes from the matched tumor genomes. We identified 438,872 somatically acquired single-nucleotide substitutions across the AA-UTUC genome, including 201,192 substitutions within genes and introns and 237,680 intergenic variants (table S3). The accuracy of the WGS data was assessed by Sanger sequencing of 250 randomly selected somatic variants (50 coding, 100 intragenic, 100 intergenic). Of these, ~98% were confirmed as genuine somatic mutations (tables S4 to S6).

The WGS analysis revealed an average mutation rate of 150 mutations/Mb of DNA in AA-UTUC, which is higher than previously reported mutation rates for smoking-associated lung cancer (8 mutations/Mb) (3) or UV radiation-associated melanoma (111 mutations/Mb) (4). Thus, among documented group 1 carcinogens, AA may exert one of the strongest known mutational pressures on the human genome (Fig. 1A).

To confirm this strikingly high AA mutational burden, we sequenced the exomes of an additional eight matched pairs of AA-UTUC. We generated an average of ~8.8 Gb of reads for each sample, yielding an average coverage of 35X in targeted regions, with 60.3% of targeted bases represented by at least 20 sequence reads (table S7). Somatic mutations were identified using bioinformatic analysis similar to the WGS analysis. We identified 9933 exonic somatic substitutions in 6415 different genes in the nine AA-UTUCs, including 8144 missense, 853 nonsense alterations, and 936 splice-site mutations. A thousand of these exonic mutations (at least 100 candidate somatic coding mutations per tumor) were selected for Sanger sequencing verification. Thirty mutations could not be tested because of polymerase chain reaction (PCR) failure, and of the remaining 970 mutations, 946 were confirmed by Sanger sequencing (97.5%, table S8). On average, 1118 nonsynonymous somatic mutations were identified for each AA-UTUC exome (904 missense; 94 nonsense, 104 splice site). We then compared the AA-UTUC mutation rate against those observed in other cancers, including (i) UV-exposed melanomas (n = 7) (20), (ii) smoking-associated lung cancers (n = 12) (21), (iii) liver fluke–induced bile duct cancers (n = 8) (2), and (iv) Helicobacter pylori–associated gastric cancers (n = 15) (1). We chose these four cancer types because they are also caused by group 1 carcinogens. Again, AA-UTUC exhibited the highest rate of exonic mutations (Fig. 1B and fig. S1). None of the AA-UTUC cases exhibited evidence of microsatellite instability (MSI; table S1), indicating that the high mutational load in AA-UTUC is not due to a classical MSI mechanism.

Fig. 1. Mutation counts in AA-UTUCs and other group 1 carcinogen-associated cancers. (A) Total numbers of single-nucleotide somatic mutations in the genomes of AA-UTUC, UV-associated melanoma, and tobacco-associated lung cancer. (B) Superimposed individual tumor data points for the total numbers of nonsynonymous single-nucleotide mutations in AA-UTUC, UV-associated melanoma, tobacco-associated lung cancer, Opisthorchis viverrini (OV)–associated cholangiocarcinoma (CCA), and H. pylori–associated gastric cancer.
Numerous genes were affected by somatic mutations in two or more of the nine tumors, an unsurprising result given the exceedingly high AA-UTUC mutation rate. Frequently mutated genes in AA-UTUC included TP53 (mutated in five of nine tumors) (13, 15) and the histone H3K27 demethylase KDM6A (eight of nine tumors). We also observed frequent mutations in other chromatin modifier genes such as ARID1A and SETX (Table 1). We did not observe mutations in the oncogenes HRAS or FGFR3, previously described to be mutated in AA-UTUC, which may be due to the relatively low mutation rates reported for these genes (4.7 and 4%, respectively (13)).

**AA induces a characteristic mutational signature in cancer**

The mutational effects of AA are thought to involve AA metabolites that bind to the amino groups of purine bases (A and G) to form DNA adducts (22, 23). In AA-UTUC, we observed a strikingly high proportion of somatic A:T to T:A mutations—a mutation class that constitutes only a minor proportion in other cancers, including *H. pylori*-associated gastric cancer and liver fluke–induced bile duct cancer (Fig. 2A) (1, 2, 4, 24–26). We did not observe high rates of mutations at guanine bases, suggesting that DNA adducts formed by AA primarily affect adenine sites in vivo.

To better understand AA mutagenesis, we examined the sequence contexts of the A:T to T:A mutants. Collectively analyzing 267,192 A-to-T transversions in the genome sequence of tumor 9T, we observed a dramatic overrepresentation of cytosines and thymines immediate′ to mutated adenines (that is, [C|T]A) and overrepresentation of guanines 3′ to mutated adenines (that is, AG) (Fig. 2B). Indeed, TAG and CAG were the most prevalent among the 16 possible A-to-T transversion trinucleotide groups (Fig. 2B, and tables S9 and S10). We did not observe high rates of mutations at guanine bases, suggesting that DNA adducts formed by AA primarily affect adenine sites in vivo.

When we compared the total number of somatic mutations occurring on nontranscribed versus transcribed strands, we discovered over twice as many A-to-T transversions on the nontranscribed strand (Fig. 2C). This strand bias suggests that AA-UTUC is similar to some other cancer types (20, 27) where lesions occurring on transcribed strands may be identified and corrected by transcription coupled repair.

Finally, 7.4% of A:T to T:A transversions across the nine AA-UTUC exomes occurred at CAG trinucleotides at 3′ splice sites. The AA-UTUC mutations at 3′ splice-site CAGs occurred at significantly higher prevalence than expected by chance (Fig. 3A and table S15; \( P = 6.6 \times 10^{-9} \) to 0.015 by hypergeometric tests). Similar 3′ splice-site enrichments were not observed in non–AA-associated cancers (Fig. 3A and table S15; \( P = 0.21 \) and 0.76, respectively, by hypergeometric tests). These findings suggest that splice-site mutations may contribute to the pathogenesis of AA-UTUC.

**AA is sufficient to recapitulate Aristolochia-induced mutagenicity and nephrotoxicity**

Human consumption of AA typically occurs via herbal remedies (9) containing AA and mixtures of other natural products. We sought to experimentally verify that AA alone, as a purified isolated compound, is sufficient to cause both renal pathology and the mutational signature observed in AA-UTUC primary tumors. We first established in vitro models by chronic treatment of human renal proximal tubular cells (HK2) with sublethal levels of AA (10 μM) for 6 months. Eighty-five percent of HK2 cells underwent apoptosis and/or necrosis upon AA treatment, but subsequently, single clones developed and proliferated.

**Table 1. The top 15 recurrent mutated genes in AA-UTUC.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>3T</th>
<th>6T</th>
<th>9T</th>
<th>10T</th>
<th>13T</th>
<th>20T</th>
<th>79T</th>
<th>80T</th>
<th>100T</th>
<th>Total cases</th>
<th>Exon numbers</th>
<th>Coding region sizes (bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDM6A*</td>
<td>8/9</td>
<td>29</td>
<td>4,203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRRK2</td>
<td>7/9</td>
<td>51</td>
<td>7,581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCHS2</td>
<td>7/9</td>
<td>25</td>
<td>8,748</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USH2A</td>
<td>7/9</td>
<td>72</td>
<td>15,615</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCN1A</td>
<td>6/9</td>
<td>26</td>
<td>6,027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADAMTS1L1</td>
<td>5/9</td>
<td>29</td>
<td>5,286</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATRX</td>
<td>5/9</td>
<td>35</td>
<td>7,476</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNAH9</td>
<td>5/9</td>
<td>69</td>
<td>13,458</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYOSC</td>
<td>5/9</td>
<td>41</td>
<td>5,226</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP53</td>
<td>5/9</td>
<td>11</td>
<td>1,023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARID1A*</td>
<td>3/9</td>
<td>20</td>
<td>6,855</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHD6</td>
<td>4/9</td>
<td>20</td>
<td>8,145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDH10</td>
<td>4/9</td>
<td>12</td>
<td>2,364</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREBBP</td>
<td>4/9</td>
<td>31</td>
<td>7,326</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SETX*</td>
<td>4/9</td>
<td>26</td>
<td>8,031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Chromatin modifier.
We randomly selected two independent AA-treated HK2 clones and their untreated parental clones for exome sequencing (31X to 34X coverage) and analysis (table S19). Sanger validation of somatic variants yielded a specificity of 95.7% (89 of 93 mutations; table S20). We identified 219 and 162 somatic substitutions in the two clones (293 missense, 31 nonsense, and 57 splice-site mutations). The mutational signature in the in vitro AA-treated HK2 clones was similar to that observed in primary AA-UTUCs, including a predominance of A:T to T:A transversions on nontranscribed strands (Fig. 4A), a [C|T]ΔG mutational motif (Fig. 4B), and an enrichment of splice-site mutations (Fig. 3A). These in vitro results demonstrated that AA alone is sufficient to recapitulate the repertoire of genomic hallmarks observed in human AA-UTUC.

We also studied AA toxicity in vivo. Three-day oral dosing of C57BL/6 mice with AA (50 mg/kg) induced a range of renal pathologies over a 90-day posttreatment period. These renal abnormalities were highly similar to previously reported *Aristolochia*-associated pathologies in humans (9, 30), including organ atrophy, tubular necrosis, and lymphocytic infiltrates in the deep cortex, outer medulla, and medullary...
In all the AA-UTUCs rank higher than in all the non AA-UTUCs. The Wilcoxon rank-sum tests for each gene. The in AA-UTUC and non of all CAGs (both mutated and wild-type) occurring at 3

Genomic screening reveals AA-like mutational signatures using the following criteria: (i) A > T is the most frequent in AA-UTUC and non–AA-UTUC quantified by quantitative RT-PCR. Results were analyzed by one-sided Wilcoxon rank-sum tests for each gene. The P values are identical because, for every gene, the mRNA levels in all the AA-UTUCs rank higher than in all the non–AA-UTUCs. (C) A heterozygous splice-site mutation resulted in skipping of MBOAT7 exon 3 in AA-UTUC. Bridging reads, confirming the exon skipping, are shown. (D) Verification of MBOAT7 exon skipping by RT-PCR. Arrows schematically indicate primer locations.

of all CAGs (both mutated and wild-type) occurring at 3’ splice sites. (B) mRNA levels of NMD machinery genes in AA-UTUC and non–AA-UTUC quantified by quantitative RT-PCR. Results were analyzed by one-sided Wilcoxon rank-sum tests for each gene. The P values are identical because, for every gene, the mRNA levels in all the AA-UTUCs rank higher than in all the non–AA-UTUCs. (C) A heterozygous splice-site mutation resulted in skipping of MBOAT7 exon 3 in AA-UTUC. Bridging reads, confirming the exon skipping, are shown. (D) Verification of MBOAT7 exon skipping by RT-PCR. Arrows schematically indicate primer locations.

DISCUSSION
Exposures to exogenous carcinogens are known to leave their imprints as mutational patterns in cancer genomes. Characterizing these signatures not only sheds light on the molecular mechanism of carcinogenesis but may also provide a genomic tool to detect a carcinogen’s involvement in cancers not previously known to be linked to the carcinogen. To date, mutational signatures in cancer have been reported for some group 1 carcinogens, including UV light, liver flukes, tobacco, and H. pylori (1–4, 27, 32). Here, we identified a highly distinctive genome-wide mutational signature associated with AA, a group 1 carcinogen associated with UTUC. Compared with other group 1 carcinogens studied to date on a genome-wide scale, AA causes the highest rate of mutations. The high mutation load observed in AA-UTUC is comparable to those observed in “hypermutated” cancers caused by mismatch repair defects (33) but lower than those observed in “ultramutated” cancers caused by inactivation of the DNA polymerases POLE or POLD1 (34) (table S21). The high level of de novo mutations identified in AA-UTUC may explain the epidemiological observation that AA-UTUC patients tend to be younger and have higher contralateral recurrence rates (16) than non–AA-UTUC patients. Beyond the sheer abundance of mutations, AA mutational events were characterized by a predominance of A:T to T:A transversions in an A[C|T]AGG context. Furthermore, AA mutations affected CAG trinucleotides at 3’ splice sites more often than expected by chance, suggesting a role for splice-site mutations in UTUC pathogenesis. This is supported by our demonstration of up-regulation of the NMD machinery and altered splicing patterns in AA-UTUC. This process, coupled with specific pro-oncogenic mutations in cancer-related driver genes (Table 1), is likely to contribute to AA-UTUC carcinogenesis. In the nine AA-UTUCs, we identified numerous genes mutated at frequencies comparable to TP53, the gene previously reported to be highly mutated in AA-UTUC (15). The most frequently mutated gene in AA-UTUCs rank higher than in all the non AA-UTUCs. The Wilcoxon rank-sum tests for each gene. The in AA-UTUC and non–AA-UTUC quantified by quantitative RT-PCR. Results were analyzed by one-sided Wilcoxon rank-sum tests for each gene. The P values are identical because, for every gene, the mRNA levels in all the AA-UTUCs rank higher than in all the non–AA-UTUCs. (C) A heterozygous splice-site mutation resulted in skipping of MBOAT7 exon 3 in AA-UTUC. Bridging reads, confirming the exon skipping, are shown. (D) Verification of MBOAT7 exon skipping by RT-PCR. Arrows schematically indicate primer locations.

of all CAGs (both mutated and wild-type) occurring at 3’ splice sites. (B) mRNA levels of NMD machinery genes in AA-UTUC and non–AA-UTUC quantified by quantitative RT-PCR. Results were analyzed by one-sided Wilcoxon rank-sum tests for each gene. The P values are identical because, for every gene, the mRNA levels in all the AA-UTUCs rank higher than in all the non–AA-UTUCs. (C) A heterozygous splice-site mutation resulted in skipping of MBOAT7 exon 3 in AA-UTUC. Bridging reads, confirming the exon skipping, are shown. (D) Verification of MBOAT7 exon skipping by RT-PCR. Arrows schematically indicate primer locations.

DISCUSSION
Exposures to exogenous carcinogens are known to leave their imprints as mutational patterns in cancer genomes. Characterizing these signatures not only sheds light on the molecular mechanism of carcinogenesis but may also provide a genomic tool to detect a carcinogen’s involvement in cancers not previously known to be linked to the carcinogen. To date, mutational signatures in cancer have been reported for some group 1 carcinogens, including UV light, liver flukes, tobacco, and H. pylori (1–4, 27, 32). Here, we identified a highly distinctive genome-wide mutational signature associated with AA, a group 1 carcinogen associated with UTUC. Compared with other group 1 carcinogens studied to date on a genome-wide scale, AA causes the highest rate of mutations. The high mutation load observed in AA-UTUC is comparable to those observed in “hypermutated” cancers caused by mismatch repair defects (33) but lower than those observed in “ultramutated” cancers caused by inactivation of the DNA polymerases POLE or POLD1 (34) (table S21). The high level of de novo mutations identified in AA-UTUC may explain the epidemiological observation that AA-UTUC patients tend to be younger and have higher contralateral recurrence rates (16) than non–AA-UTUC patients. Beyond the sheer abundance of mutations, AA mutational events were characterized by a predominance of A:T to T:A transversions in an A[C|T]AGG context. Furthermore, AA mutations affected CAG trinucleotides at 3’ splice sites more often than expected by chance, suggesting a role for splice-site mutations in UTUC pathogenesis. This is supported by our demonstration of up-regulation of the NMD machinery and altered splicing patterns in AA-UTUC. This process, coupled with specific pro-oncogenic mutations in cancer-related driver genes (Table 1), is likely to contribute to AA-UTUC carcinogenesis. In the nine AA-UTUCs, we identified numerous genes mutated at frequencies comparable to TP53, the gene previously reported to be highly mutated in AA-UTUC (15). The most frequently mutated gene in
our cohort of AA-UTUCs was KDM6A, a histone demethylase gene. KDM6A mutations have been previously observed in different cancer types, but most notably in bladder cancer, a malignancy of transitional cells that are also predominantly found in AA-UTUC (25). Besides KDM6A, we also observed mutations in other chromatin modifying genes in AA-UTUC (Table 1). To date, very little is known about how these genes, which have been recently found to be frequently mutated in cancer, contribute to tumorigenesis (35). Thus, further functional studies of these genes to elucidate their tumorigenic roles are warranted.

Several cellular pathways exist for repairing DNA lesions caused by carcinogens (36). Of these, nucleotide excision repair (NER) serves as the predominant DNA repair pathway for repairing bulky DNA adducts, involving both global genomic NER (GG-NER) and transcription-coupled NER (TC-NER) (37, 38). The strand bias that we observed in this study suggests that AA-adducts occurring on the transcribed strand may be removed and corrected by TC-NER, but not by GG-NER. AA-associated adducts have been shown to persist in human tissues for many years (39). This is in accordance with recent evidence demonstrating that AA-associated adducts can escape GG-NER because they are not detected by the XPC/RAD23B recognition complex, the initial activating step for GG-NER (40). Our data suggest that whereas TC-NER may function in some part to correct a proportion of AA-adducts, it appears to be overwhelmed by the sheer abundance of somatic mutations caused by AA exposure.

Kidneys and livers play essential roles in drug and toxin metabolism. In the case of AA, cytosolic NAD(P)H:quinone oxidoreductases expressed by these two organs can activate AA by nitroreduction to N-hydroxyaristolactam, forming a cyclic N-acylnitrenium ion as the ultimate carcinogenic species binding DNA (41). AA-DNA adducts have been detected in the stomach, kidney, urinary tract, bladder, and liver (11), suggesting a plausible carcinogenic role of AA in these organs. Given the highly distinctive genome-wide mutational signature of AA, we tested whether the AA signature could be used as a "molecular fingerprint" to screen for the potential involvement of AA in other cancers besides UTUC. Specifically, we interrogated cancer genome data for 93 HCCs, and found evidence for AA-like mutational signatures in 11 cases. In these HCCs, it is possible that AA-induced mutations may have collaborated with other oncogenic insults, such as HBV infection, to cause the development of cancer—whether these interactions are additive or synergistic warrants further investigation.

Our study has certain limitations. First, the comparisons of mutation rates in UV-, tobacco-, and AA-associated cancers were performed by comparing our rates for AA-associated cancer to published rates for the other cancers (3, 4), rather than reanalyzing the raw sequencing reads using identical variant detection pipelines. Nevertheless, our sequencing data and analysis pipeline are highly similar to those used by these other studies, and for the same tumor type, our pipeline’s
Recent studies have proposed a strong dose-effect relationship between AE exposure, nephrotoxicity, and eventual cancer over a long latency period (43). In humans, the most likely cause for these exposures is prolonged herbal intake resulting in chronic carcinogenic exposure to AA. Because little is known about the specific concentrations of AA in different herbal remedies, a more reliable method of determining extent of AA exposure in individuals involves the direct detection of AA-DNA adducts in target tissues. Different techniques including 32P-postlabeling (44) and high-performance liquid chromatography coupled with mass spectrometry (45) have been used to detect such adducts. However, these approaches have not been developed systematically as a routine diagnostic tool for detecting AA exposure because the sensitivity of these assays is highly dependent on the specific dosage, duration of exposure, and persistence of AA-DNA adducts at the time of measurement. By contrast, although cancer genome sequencing does not detect AA-DNA adducts per se, such genomic information carries the potential of reflecting the long-term cumulative mutational consequences of AA exposure. Essentially, our findings provide a proof of concept that it is technically feasible to use mutational signatures associated with a carcinogen to screen for its involvement in cancers not previously known to be caused by that carcinogen. This general approach could be further extended to all group 1 carcinogens to investigate their involvement in any cancer type for which high-throughput sequencing data are available. Moreover, it may also be possible to extend such "mutational fingerprinting" to noncancer tissues such as blood cells, which are easily accessible for screening. This general strategy may thus play an important role in the field of cancer prevention as a genomic strategy to detect early carcinogen exposures.

Nearly 80% of the world population relies on traditional medicines for their primary health care needs (46). Despite public health warnings regarding the safety of botanical products and dietary supplements containing AA and bans on the use of such products in several countries, almost 20 years later potential sources of AA remain available. For example, certain AA-containing products are still permitted under the supervision of practitioners of Chinese medicine (43), and products containing AA are still easily available worldwide. Other sources of AA include contamination of wheat flour by the seeds of Aristolochia species, as reported in the Balkans (49). Frequent intake of AAs, sometimes at very high concentrations, has been reported in studies from China, Taiwan, Hong Kong, Japan, and some Western countries (50). Thus, the strikingly high mutational burden and toxicity of AA as demonstrated in this study and another independent report (51) highlights the importance of knowing the contents of herbal products and of greater public awareness of the fact that not all derivatives from "natural" plants are safe.

MATERIALS AND METHODS

Study design

Supplementary Materials and Methods provides details of the study design.

Clinical samples and information

Tissue samples and clinical information on the subjects with AA-UTUC were obtained from the Chang Gung Memorial Hospital in Taiwan. Written consent was obtained from each subject, and the research protocol was approved by the Human Research Ethics Committee of the Chang Gung Memorial Hospital.
Sequencing

The genomes of the AA-UTUCs and the matched nonmalignant tissue sample were sequenced on an Illumina HiSeq 2000 as paired-end 76-bp reads. The exomes of nine AA-UTUCs and matched normal adjacent tissues, the HK2 cell line (from American Type Culture Collection), and two AA-treated clones of HK2 cells were sequenced on an Illumina GAIIx sequencer (2 × 76-bp reads). Read pairs were aligned to the reference human genome (hg19) using Burrows-Wheeler Aligner. Somatic single-nucleotide mutations were identified according to their presence in the tumor genome and absence from the corresponding normal genome as described in the text and Supplementary Materials. RNA sequencing was carried out on an Illumina HiSeq 2000 and analyzed as described in Supplementary Materials and Methods.

Analysis of NMD genes

Supplementary Materials and Methods provides details of the analysis of expression levels of nonsense-mediated decay genes.

Induction of AA-associated nephropathy in C57BL/6 mice

C57BL/6 mice were treated with AA (50 mg/kg) for 3 days and sacrificed on days 10, 30, and 90 to monitor AA-induced nephropathy. The procedures for the present study were approved by the Animal Committee under Singapore Health Science (SingHealth), and all animals were treated according to the guidelines for animal experimentation of SingHealth (Institutional Animal Care and Use Committee protocol #2012/SHS/773).

Statistical and bioinformatic analyses

Supplementary Materials and Methods provides details of bioinformatic and statistical analyses.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/5/197/197ra101/DC1 Materials and Methods

Fig. S1. Frequency of mutations in carcinogen-induced cancers.

Fig. S2. Single-nucleotide somatic mutations in 16 possible sequence contexts for A>T transitions in the AA-UTUC whole genome.

Fig. S3. Single-nucleotide somatic mutations in 16 possible sequence contexts for A>T transitions in the exomes of nine AA-UTUCs.

Fig. S4. Systematic up-regulation of NMD gene transcripts in AA-UTUC compared to adjacent normal tissue.

Fig. S5. A heterozygous 3′ splice-site mutation results in skipping of RFc2 exon 10 in AA-UTUC. Fig. S6. Strong association between CAG/CTG mutations at 3′ splice sites and altered splicing.

Fig. S7. Further details of the in vivo model of AA-induced damage.

Fig. S8. Superimposed individual tumor data points for the total nonsynonymous single-nucleotide variants and each of the separate mutation types in AA-HCCs and non-AA-HCCs.

Fig. S9. Nineteen HCCs exhibiting a “weak” AA mutational signature.

Fig. S10. Representation of 3′ splice-site CAGs.

Table S1. Clinical characteristics of AA-UTUC patients analyzed by whole-genome and/or exome sequencing.

Table S2. Sequence analysis summary of whole genome sequencing.

Table S3. Breakdown of somatic mutations by genomic region.

Table S4. Somatic nonsynonymous substitutions in protein-coding genes of the whole genome.

Table S5. Somatic substitutions in unspliced transcript regions (transcribed, including introns and untranslated regions) of AA-UTUC.

Table S6. Somatic substitutions in the intergenic regions of AA-UTUC.

Table S7. Sequence analysis summary of nine exome-sequenced AA-UTUCs.

Table S8. Somatic nonsynonymous substitutions in protein-coding genes of nine AA-UTUCs.

Table S9. The effect of +/− one base flanking the mutated adenine or thymidine on the number of intergenic mutations in AA-UTUC.

Table S10. The effect of +/− one base flanking the mutated adenine or thymidine on the number of intergenic mutations in AA-UTUC.

Table S11. The effect of +/− two bases flanking the mutated adenine or thymidine on the number of intergenic mutations in AA-UTUC.

Table S12. The effect of +/− one base flanking the mutated TAG on the rates of unspliced transcript (transcribed regions, including introns) mutations in AA-UTUC.

Table S13. The effect of +/− one base flanking the mutated CAG on the rates of unspliced transcript (transcribed regions, including introns) mutations in AA-UTUC.

Table S14. Hypogeometric analysis for enrichment of CAG splice-site mutations in AA-UTUCs.

Table S15. AA-treated HK2 clones, and non-AA-associated cancers.

Table S16. RPKM gene expression values for 15 NMD pathway genes in the AA-UTUC and matched normal tissue.

Table S17. Identities of 3′ splice sites with CAG/CTG mutations and RPKM > 2.

Table S18. 3′ splice sites without CAG/CTG mutations for evaluating the proportion of unspliced sites associated with aberrant splicing.

Table S19. Sequence analysis summary of two exome-sequenced AA-treated HK2 clones.

Table S20. Somatic nonsynonymous substitutions in protein-coding genes of AA-treated HK2 clones.

Table S21. Comparison of mutation rates in AA-UTUC, carcinogen-induced cancers, mismatch repair-defective colorectal cancers, and POLE/POLD1 mutated colorectal cancers.

Table S22. Primer sequences.

REFERENCES AND NOTES


Aristolochia fangchi, a Chinese herbal product containing aristolochic acid.
Funding: This work was supported in part by funding from the Singapore National Medical Research Council (NMRC/STAR/0006/2009), the Singapore Millennium Foundation, the Lee Foundation, the Singapore National Cancer Centre Research Fund, The Verdant Foundation, the Duke-NUS Graduate Medical School, the Cancer Science Institute, Singapore, the Chang Gung Memorial Hospital (CMRPG3/70471-3), the Taiwan National Science Council (98-2314-B-182A-057-MY3), and Wellcome Trust (grant reference 098051). W.Y. and L.D.L. are the recipients of the NUS Graduate School for Integrative Sciences and Engineering Scholarship, Singapore. S.G.R. received startup funds from the Singapore Ministry of Health and Agency for Science, Technology, and Research via Duke-NUS Graduate Medical School. Author contributions: S.L.P., J.R.M., S.-T.P., S.G.R., P.T., and B.T.T. conceived the study. S.-T.P., S.G.R., P.T., and B.T.T. directed the study. W.-H.W., Y.-H.C., C.-K.C., K.-J.Y., K.-F.N., C.-F.W., and C.-L.H. were involved in the specimen collection, pathological reviews, and clinical data collections. S.L.P., E.Y.S., S.C.C., and R.M.B. were involved in establishment of the in vitro and in vivo models of AA-induced damage. J.R.M., W.Y., K.K.H., W.K.L., P.G., Y.L., I.C., M.R.S., P.A.F., W.-K.S., and S.G.R. performed the bioinformatics data analysis. S.L.P., H.L.H., A.G., S.T.T., D.H., L.D.L., M.L.N., W.C.-o., and C.K.O. performed sequencing and quantitative RT-PCR experiments. S.L.P., J.R.M., S.-T.P., S.G.R., P.T., and B.T.T. wrote the manuscript, with the assistance and final approval of all authors. Competing interests: The authors declare that they have no competing interests. Data and materials availability: The data for this study have been deposited in the European Nucleotide Archive under accession number PRJEB4138. Correspondence and requests for materials should be addressed to B.T.T., P.T., S.G.R., or S.-T.P.

Submitted 7 March 2013
Accepted 5 July 2013
Published 7 August 2013
10.1126/scitranslmed.3006086

Editor's Summary

Carcinogen AAAlert

Aristolochic acid (AA) is a natural compound derived from plants in the Aristolochia genus. For centuries, Aristolochia has been used throughout Asia to treat a variety of ailments as a component of traditional Chinese medicine. In recent years, however, a more sinister side of this herb has come to light when it was linked to kidney damage and cancers of the urinary tract. Now, two studies by Poon et al. and Hoang et al. present a "molecular signature" of AA-induced DNA damage, which helps to explain the mutagenic effects of AA and may also be useful as a way to detect unsuspected AA exposure as a cause of cancer.

The molecular signature seen in AA-associated tumors is characterized by a predominance of A:T-to-T:A transversions, a relatively unusual type of mutation that is infrequently seen in other types of cancer, including those caused by other carcinogens. These mutations concentrate at splice sites, causing the inappropriate inclusion or exclusion of entire exons in the resulting mRNA. The overall mutation rate is another notable feature of AA-associated cancers because it is several times higher than the rate of mutations caused by other carcinogens such as tobacco and ultraviolet light. In both studies, the authors also used the molecular signature to discover that AA was a likely cause of tumors previously attributed to other carcinogens. In one case, a urinary tract cancer that had been attributed to smoking and, in the other case, a liver cancer previously attributed to a chronic hepatitis infection were both identified as having the telltale signature of AA mutagenesis.

The identification of a specific molecular signature for AA has both clinical and public health implications. For individual patients, the molecular signature could help physicians identify which tumors were caused by AA. Although this information cannot yet be used to optimize the treatment of individual patients, those who are diagnosed with AA-associated cancers could be monitored more closely for the appearance of additional tumors. Meanwhile, a better understanding of the mutagenic effects of AA should also help to strengthen public health efforts to decrease exposure to this carcinogenic herb.
A complete electronic version of this article and other services, including high-resolution figures, can be found at:
http://stm.sciencemag.org/content/5/197/197ra101.full.html

Supplementary Material can be found in the online version of this article at:
http://stm.sciencemag.org/content/suppl/2013/08/05/5.197.197ra101.DC1.html

Related Resources for this article can be found online at:
http://stm.sciencemag.org/content/scitransmed/2/32/32ra35.full.html
http://stm.sciencemag.org/content/scitransmed/5/197/197fs31.full.html
http://stm.sciencemag.org/content/scitransmed/2/19/19ra13.full.html
http://stm.sciencemag.org/content/scitransmed/5/197/197ra102.full.html
http://stm.sciencemag.org/content
http://stm.sciencemag.org/content/scitransmed/6/221/221ra15.full.html
http://www.sciencemag.org/content/sci/344/6180/146.full.html
http://www.sciencemag.org/content/sci/346/6215/1291.full.html
http://www.sciencemag.org/content/sci/346/6215/1373.full.html
http://www.sciencemag.org/content/sci/347/6217/12.full.html
http://www.sciencemag.org/content/sci/347/6217/81.full.html
http://www.sciencemag.org/content/sci/347/6217/78.full.html
http://www.sciencemag.org/content/sci/347/6225/1006.full.html

Information about obtaining reprints of this article or about obtaining permission to reproduce this article in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl